Original Article

Extensive Deep Belief Nets with Restricted Boltzmann Machine Using MapReduce Framework

Pandiganesh S.* and J.C. Miraclin Joyce Pamila

CSE Department, Government College of Technology, Coimbatore, India

Address for Correspondence

CSE Department,
Government College of Technology,
Coimbatore, India.
E-mail: pgramesh007@gmail.com

ABSTRACT

Big data is a collection of data sets which is used to describe the exponential growth and availability of both ordered and amorphous data. It is difficult to process big data using traditional data processing applications. In many practical problems, deep learning is one of the machine learning algorithms that has received great popularity in both academia and industry due to its high-level abstractions in data by using model architectures. In deep learning, Deep Belief Nets are stacks for restricted Boltzmann Machine and it is the most important deep layered (multi-layer) architecture. A restricted Boltzmann Machine is the energy based models for pre-training and followed by fine-tuning the whole net using back propagation. This mainly involves the process of classification of data. This paper contains a survey about Deep Belief Nets (DBNs) which is composed of multiple layers of latent variables along with stacked restricted Boltzmann Machine and describes big data processing. The survey concludes that the learning of Deep Belief Nets (DBNs) has attracted widespread attention due to their efficient performance in various applications.

Keywords: Big data, Map reduce, Deep belief nets, Restricted boltzmann machine.

INTRODUCTION

In recent times, the technologies like Internet of Things have created a scenario, where large amount of data (typically known as Big Data) is generated every passing moment. In monitoring this data, there lie many opportunities for various organizations. The organizations and companies handling the big data have to come up with improved capabilities and methodologies to process the data, which in turn will increase their potential operating margins. These companies normally sit in the middle of this huge information which enables them to keep track of the consumer preferences and also how the data flows among the products and services. The technologies deployed by the organizations increase the value of data significantly by making information transparent. This in turn makes the data readily available to be used in next generation of products and services. Considering the above trends this paper focuses on the deep learning and MapReduce mechanisms to monitor the big data. Deep Learning is one of the prominent technology utilized in the field of Machine learning and Data Science. The Deep Learning has been rented from the field of Artificial Intelligence, in particular from Neural Network technology. It shows promising results by solving large and high dimensional problems, feature extraction and classification. The Deep learning mechanism provides solutions for multiple disciplines that are prone to Big Data problems. MapReduce is a programming paradigm for processing and generating large data sets. A map job takes a set of information and converts it into another set of information where the elements are broken into key/value pairs. The reduce job uses the output generated from the map job and combines that information into smaller set of records. In MapReduce function the reduce job is always performed after the map job.

The paper is structured as follows. Section II provides a brief background on Deep Belief Nets. Section III describes methods used by Restricted Boltzmann Machine to manage big data. Section IV provides a basic background on MapReduce. Section V comprises of the conclusion.

Deep belief nets

The Deep Belief Nets (DBNs) is one of the recent works, which is possible to learn multiple layers of latent variables composed of stochastic variables[1-3]. DBNs can learn probabilistically to modernize its inputs. The layers in DBN act as the feature detectors for inputs and can be further trained to perform classification. DBN is also knows as multilayer generative model and its features are slightly produces a neural net to train randomly with initialized weights and biases[4]. The deep belief nets have been successfully used for many other techniques. Some recent works on DBNs include handwritten digits, Face images, Information retrieval, and Text documents. In this paper the various applications and techniques that are achieved by DBN are discussed. Figure 1 illustrates the Deep Belief Nets with Restricted Boltzmann Machine.

Complementary priors

Geoffrey E. Hinton[5] describes the fast learning algorithm for deep belief nets. He shows how to use the complementary priors in densely connected networks over many hidden layers and also states that the possibility of densely connected deep belief network one layer at a time. The implementation of complementary prior makes true posterior instead of ignoring higher layers and the tied weights are assumed. These weights are equivalent to undirected mod which is learning efficiently using Contrastive Divergence. The function of each layer has been learned and its weights are untied from the weights in higher layers. The changes in weights of higher-layer to the priors for lower layers cease to be complementary, so that the true posterior distributions in lower layers are no longer factorial.

$$p(s_i=1) = \frac{1}{1+\exp\left(-bi-\sum js_j w_{ij}\right)} \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Equation 1 describes the generation of data over the net. The probability of training on unit 'i' is a logistic function, j is the immediate ancestors and w_{ij} is weights on the directed connections from the ancestors, where b_i is the bias of unit 'i'. (See figure 1.)

Greedy layer-wise training

The DBN had been trained one layer at a time by using Greedy Layer-wise training algorithm[6]. It first trains an RBM which takes the empirical data as the input and processes it. This process gives rise to an empirical distribution over first layer of the posterior is sampled from the distribution of empirical data. So it is defined that the First level DBN is an RBM. In greedy layer-wise training it defines that after training the top-level of RBM at *l*-level DBN changes the interpretation of the RBM parameters to insert into (*l*+1) level of DBN. The distribution is RBM says that the next level from the top-level is kept as part of the DBN generative model. In the RBM, between two layers (*l*-1) and *l* the distribution is defined in terms on the parameters of RBM, whereas in the DBN it is defined in terms of the parameters of the upper layers. The input model of the RBM does not correspond to the distribution in the DBN except when the RBM is the top layer of the DBN.

Top-level model

The top-level model for DBN deals with 3D Object recognition[7]. The deep belief net for 3D Object classification has no prior knowledge in spatial structure to monitor all the variations by lighting and viewpoint, so it requires more parameters. It is not clear that the DBNs perform 3D Object classification when it is compared with shallow techniques. The top level model is trained using a combination of generative and discriminative gradients. They are using NORB database which is designed for object recognition tasks that requires generalization under varying lighting conditions. The top-only model considered for a DBN is an RBM contains two types of observed units i.e. one for the label and another for the penultimate feature vector. The energy function is considered both penultimate layer v_i, hidden layer h_j and label unit l_k, it is defined as the Boltzmann machine with three-way cliques.

$$E(v,h,l) = -\sum_{i,j,k} w_{i,j,k} v_i h_j l_k \quad \text{.................(2)}$$

The generative modeling in object recognition is an unsupervised greedy generative learning which is to extract an image representation that supports more accurate than the raw pixel representation.

Restricted boltzmann machine

The Restricted Boltzmann Machine comprises of neural networks that perform unsupervised learning[8]. In this implementation the neural network learns about the input data based on the training sets containing input factors. One of the applications for unsupervised learning performed by neural networks in Restricted Boltzmann Machines is data extraction. In this process, the meaningful features of data are extracted and simultaneously it makes it easier to protect the information about the data. It leverages the availability of unlabeled data. Figure 2 gives an undirected graphical model of an RBM contains the fine distribution of the input factor and also it contains visible layer and hidden layer with binary units. The input layer and hidden layer with an array of connection having assigned weights between the input and hidden neurons and there is no connection between the same layer.

The distribution consists of the energy function given by the equation 3.

$$E(x,h) = -h^TW_x-c^Tx-b^Th \quad \text{....................(3)}$$

The energy function contains the visible layer x, hidden layer h. The vectors b and c are the biases of the visible layer and the hidden layer and W is the matrix which is fully connected weight between two layers. Thus the distribution function consists of the probability of energy function given by equation 4.

$$p(x,h) = \exp(-E(x,h))/Z \quad \text{.....................(4)}$$

Where, Z is the partition function.

Contrastive divergence

Hinton proposed the Contrastive Divergence learning algorithm. The idea of contrastive divergence is to replace the expectation by a point estimate at single data point ($\bar{x}$). The distribution gives by Gibbs sampling specifically for RBM. The performing of sampling is efficient because condition in one layer and the entire element in other layer are dependent and sample all values in one layer are parallel given the value of opposite layer an alternate between the layers.

Continuous restricted Boltzmann machine

The diffusion network with symmetric connection for the continuous restricted Boltzmann machine[9] forms a constrained diffusion network by applying minimizing contrastive divergence rule. The continuous-valued of stochastic units that the CRBM offers improved modeling ability with both artificial and real continuous data. The CRBM models the continuous data with the simplified MCD rule. It is used to develop stochastic behavior and gives response to training which is ranging from binary stochastic to deterministic. So the modeling flexibility exceeds the probabilistic models with more binary units. It is clear that the artificial data designed to investigate the CRBM forms a generative model.

Discriminative restricted boltzmann machine

The Discriminative Restricted Boltzmann Machine describes that the RBMs are trained specifically to be food for classification models[10]. This model is appropriate and it can be advantageous to optimize directly to the input vector (x_i) and target class (y_i).

$$L_{disc}(D_{train}) = \sum_{i=1}^{|D_{train}|} \log p(y_i \mid x_i) \quad \ldots\ldots\ldots\ldots(5)$$

RBMs are trained according to L_{disc} discriminative RBMs. Since RBMs has enough hidden units and are universal approximators for binary inputs. The DRBMs are universal approximators of conditional distributions with binary inputs. It can be trained by Contrastive Divergence with conditional RBMs. This discriminative approach is used previously for fine-tuning the top RBM of a DBN.

Overview of MapReduce

MapReduce is a programming paradigm. It is a software framework which is easy to write applications that process large amount of data in parallel on a large cluster of commodity hardware in fault-tolerant manner[11]. Figure 3 shows an overview of MapReduce framework. MapReduce is the combination of two components, i.e. map function and reduce function. The map function having a list of key/value pair and the output is a list of intermediate key/value pair. Reduce function takes the intermediate key/value pair and associated with the same key and produce the list of key/value pair. The output is a final output of MapReduce processing. The MapReduce job divides the input data set of independent chunks that needs to be processed completely in a parallel manner. This framework sorts the output of the maps which is the input for reduce tasks. Both the input and output are stored in a file-system. The MapReduce framework takes care of scheduling tasking, monitoring them and re-executes in case of failure tasks.

Hadoop

Hadoop is one of the open source software framework for big data processing in a parallel manner on a large cluster. Hadoop[12] supports massive data storage and fast processing. It has the ability to process and store any kind of huge data. The volumes and varieties of data growing every day especially from social media. It is an open source framework which is free and use to store large quantities of data. Hadoop has three core components namely HDFS, MapReduce, and

YARN. HDFS (Hadoop Distributed File System) is a java based distributed file system that can store all kinds of data. MapReduce is a programming model which is used to process huge amount of data in parallel. YARN is a resource management framework. It is for scheduling and handling resource requests in distributed applications. The remaining paper discusses about MapReduce framework.

Master

The master keeps the data structures for each map task and reduce task. It stores the state and identity of the worker machine[13]. The intermediate file region is propagated from map tasks to reduce tasks, so it stores the locations and sizes for each completed tasks where the file region is produced by the map task. If the map task is completed it updates the location and size information is received.

Task granularity

The map phase are divided into M pieces and the reduce phase are divided into R pieces. M and R should be larger than the number of worker machines[14]. Each worker performs different tasks and improves dynamic load balancing and also the speed of recovery when the worker fails. R is constrained by the users since the output of each reduced task ends up in a separate output file.

Input and output types

The MapReduce provides supporting data in different formats[15,16]. For example with text mode as input it reads each line as a key/value pair. The key is denoted as the offset in the file and the value is the content of the line. Every input type knows how to divide itself into meaningful ranges for processing a map tasks. The set of output types produce data in different formats which is easy for user code support for the new output types.

CONCLUSION

In this paper, a Deep Belief Nets is reviewed and the methodologies of Restricted Boltzmann Machine are discussed with regard to big data processing. Then the paper gives brief introduction of MapReduce framework and the importance of hadoop in big data processing. This survey paper will help to understand the functionality of Deep Belief Nets, the implementation of Restricted Boltzmann Machine in big data processing and how neural networks are employed in processing of big data over distributed environment.

REFERENCES

1. G. Hinton *et al.*, ``Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82-97, Nov. 2012.
2. G. E. Dahl, D. Yu, L. Deng, and A. Acero, ``Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 30-42, Jan. 2012.
3. A.-R. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, ``Deep belief networks using discriminative features for phone recognition," in *Proc. IEEE ICASSP*, May 2011, pp. 5060-5063.
4. KUNLEI ZHANG, XUE-WEN CHEN, "Large-Scale Deep Belief Nets With MapReduce", Digital Object Identifier 10.1109/ACCESS.2014.2319813.
5. G. E. Hinton, S. Osindero, and Y.-W. Teh, ``A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527-1554, 2006.
6. Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, ``Greedy layer-wise training of deep networks," in *Advances in Neural*

Information Processing Systems, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 153.
7. V. Nair and G. E. Hinton, ``3D object recognition with deep belief nets,'' in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2009, pp. 1339-1347.
8. G. E. Hinton, “A Practical Guide to Training Restricted Boltzmann Machines,” in *Technical report 2010-003, Machine Learning Group, University of Toronto*, 2010.
9. H. Chen, A.F. Murray, “Continuous restricted Boltzmann Machine with an implementable training algorithm”, Image Signal Process, Vol. 150, No. 3, June 2003.
10. H. Larochelle and Y. Bengio. Classification Using Discriminative Restricted Boltzmann Machines. In ICML, pages 536–543, 2008.
11. J. Dean and S. Ghemawat, ``MapReduce: Simplified data processing on large clusters,'' *Commun. ACM*, vol. 51, no. 1, pp. 107-113, 2008.
12. C. Chu *et al.*, ``Map-reduce for machine learning on multicore,'' in *Neural Information Processing Systems*, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 281.
13. B. Panda, J. S. Herbach, S. Basu, and R. J. Bayardo, ``PLANET: Massively parallel learning of tree ensembles with MapReduce,'' *Proc. VLDB Endowment*, vol. 2, no. 2, pp. 1426-1437, 2009.
14. W. Zhao, H. Ma, and Q. He, ``Parallel *K*-means clustering based on MapReduce,'' in *Proc. Int. Conf. Cloud Comput.*, 2009, pp. 674-679.
15. Y. He *et al.*, ``MR-DBSCAN: An efficient parallel density-based clustering algorithm using MapReduce,'' in *Proc. IEEE 17th ICPADS*, Dec. 2011, pp. 473-480.
16. Y. LeCun, F. J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In CVPR, Washington, D.C., 2004.

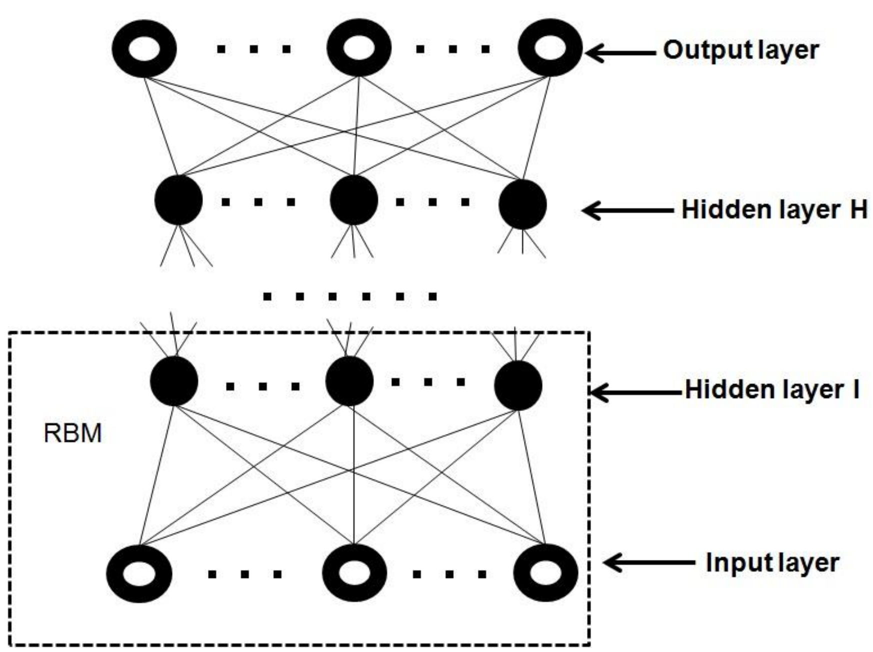

Figure 1. Deep Belief Nets with RBM

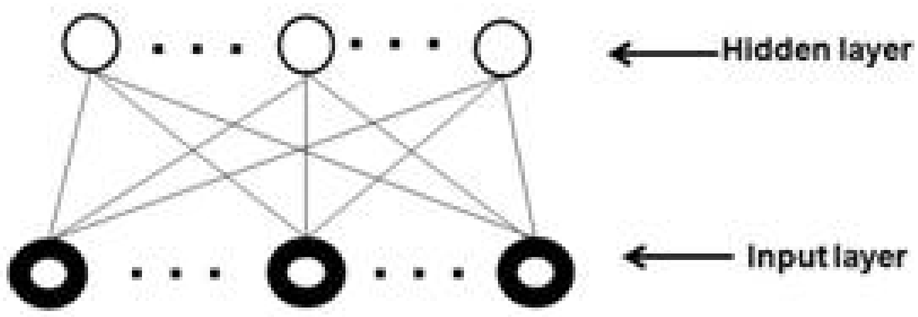

Figure 2. Restricted boltzmann machine

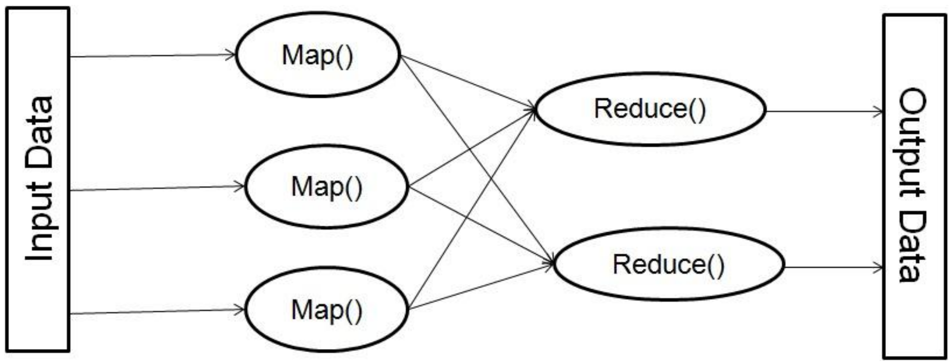

Figure 3. Map reduce